

长江水质的分析和预测

摘要

本文基于十年内长江流域水流量和水质的数据，对长江水质作出定量的综合分析，研究并探讨了水质的变化及其影响因素，对未来十年的长江水质作出预测。

对于问题 1，分别给出 4 个观测项目标准下的各水质等级观测值数量的比例，及其地域分布规律。对于水质的综合评价，我们首先以单指标最差原则得出各个河段水质的综合评价结果；而后，引入模糊评价方法，以各水质等级作为污染物浓度的模糊子集，计算各断面的污染程度对水质类别的隶属度，得出各河段水质的综合评价。

对问题 2，将干流污染问题简化为非点污染源对干流污染物浓度变化的影响问题。在此条件下根据质量守恒原理与水动力学得出一维水质模型微分方程，在非点源为线源形式时，分别给出流速不变与河道横截面积不变两种情况下浓度的解析解，继而求出各地区污染源的平均污染物浓度，进而得到两种主要污染物的主要污染源均在宜昌到岳阳地区和岳阳到九江地区的结论。

对于未来十年的水质预测，为了降低预测的复杂性，我们首先通过聚类分析将六类水质聚合为三个等级，然后采用时间序列分析和多元分析两种数理统计方法对水质进行预测。对于前者，我们尝试了指数平滑法和神经网络算法做时间序列分析，得到未来水质百分比变化。对于多元分析，我们构造简单的水箱模型作出水质等级百分比和年废水排量与年流量比值存在线性关系的猜想，并通过相关性分析验证此猜想，建立线性预测模型；同时，通过曲线拟合得出了模型中的重要因子年废水排放量和年流量的未来十年变化规律，从而预测出对应年份的各水质等级河段的比例：

预测	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
I,II,III	68.5%	66.1%	64.5%	63.9%	63.6%	63.1%	61.7%	59.4%	56.1%	52.3%
IV,V	19.7%	20.9%	21.6%	21.8%	22.0%	22.2%	22.8%	23.9%	25.4%	27.1%
劣 V	11.7%	13.1%	13.9%	14.3%	14.4%	14.7%	15.5%	16.7%	18.6%	20.6%

对于问题 4，我们根据线性预测模型的预测结果，以处理污水总量最小为目标建立线性规划模型，在满足劣 V 类水和 IV，V 类水比例的约束条件下得出每年需要处理的污水量(如下表)并分析了两类约束对最优解的影响。

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
处理量	141.5	155.8	167.1	175.8	183.5	192.2	203.5	218.1	235.1	252.8

关键字：隶属度，一维水质模型，聚类分析，时间序列指数平滑法，神经网络，相关性分析，多元分析法

长江水质的分析和预测	1
1.摘要	1
2.问题重述	2
3.问题分析	3
3.1 水质的综合评价	3
3.2 主要污染源	3
3.3 水质的预测	3
3.4 每年所需处理的污水量	3
4.符号约定	4
5.模型假设	4
6.水质综合评价，地区污染状况：对问题 1 的回答	4
6.1 总体概览	4
6.2 单项目指标评价	5
6.3 地区单项目对比评价	6
6.4 综合评价	8
7.污染源分析：对问题 2 的回答	11
8.预测模型的建立和求解：对问题 3 的回答	14
8.1 对六种类别的聚类分析：	14
8.2 时间序列分析预测：指数平滑法	15
8.3 时间序列分析预测：神经网络	16
8.4 多元分析预测：线性预测模型	16
8.4.1 年废水排放量 W 与年总流量 F 对水质的函数关系的猜想	17
8.4.2 单位流量的污水排放量 R 与各水质等级比例的函数关系确定	19
8.4.3 年污水排放量 W 的预测	21
8.4.4 水流量 F 的预测	22
8.4.5 未来十年内各等级水河段比例的趋势	23
9.模型应用：对问题 4 的解答	24
10.污染治理方略：对问题 5 的回答	25
11.模型评价	26
11.1 时间序列分析方法和多元分析方法	26
11.2 线性预测模型的局限性	26
12.模型拓展	27
13.参考文献	27

2.问题重述

根据附表对长江近两年多的水质情况做出定量的综合评价，并分析各地区水质的污染状况；研究、分析长江干流近一年多主要污染物高锰酸盐指数和氨氮的污染源；假如不采取更有效的治理措施，依照过去 10 年的主要统计数据，对长江未来水质污染的发展趋势做出预测分析，比如研究未来 10 年的情况；根据预测分析，确定每年处理污水量使得未来 10 年内每年长江干流的Ⅳ类和Ⅴ类水的比例在 20% 以内，且没有劣Ⅴ类水。

3.问题分析

3.1 水质的综合评价

我们认为，对水质的评价分为层次评价和综合评价，前者是以某一个因素作为分类变量对不同因素下的水质进行单独评价和对比；而后者是在一定的原则下，对各地区的水质用一个综合的参数来描述。对于前者，我们分别在 4 个观测项目标准的层面下分析了整个流域各采样点在各个水质等级的分布，同时在不同地区的层面上对各污染指标在各地区的分布作了统计分析。在综合评价中，当综合评价指标反映该地区水质满足饮用水要求的水平时，应该以四个水质水平中最差的一项作为该地区污染水平的度量；当此指标反映该地区各个水质因素的综合影响时，应考虑四个水质标准的某种方式下的加权平均。

3.2 主要污染源

我们认为，主要污染源是单位时间内污染物排放量最大的若干河段。因此，主要污染源的确定问题等效为污染物排放量的确定问题。各河段污染物主要来自于上流的污染物和本河段的污染物排放，而本河段污染量可由观测点的水流量、流速和污染物浓度确定。因此，在确知本河段的水流量、流速、污染物浓度和上游河段的这三个量的值的情况下，建立相应的水质变化模型描述污染物浓度随漂流距离的变化，可以求出个河段的污染物排放量。

3.3 水质的预测

水质的预测可以用两种不同的数理统计方法。一是将水质自身的历史变化视为时间序列，对此序列进行统计分析寻求其规律性，并据此对水质未来的变化作出分析预测；第二种方法是寻求影响水质的因素并通过统计分析找出两者的统计关系，通过对影响因素的预测，得出水质的预测值。我们对于这两种方法都作了尝试并分析其结果。

3.4 每年所需处理的污水量

我们认为短期内，每年污水的排放量只由当年的工业发展需求与居民生活需求决定，而与上一年的水质和治理情况无关，则根据污水排放量的预测值，并结合上题中污水排放量与水质的关系，可以计算出达到一定水质下每年所需处理的污水量。

4.符号约定

- Q_i : 干流上各观察站处水流量
- V_i : 干流上各观察站处水流速
- C_i : 干流上各观察站处水中污染物浓度
- $C_{d,i}$: 干流上相邻观察站之间地区所接纳水流平均污染物浓度
- M_i : 干流上相邻观察站之间地区单位时间排放污染物的量
- L_i : 干流上相邻观察站之间距离
- k : 污染物降解速率, 本题中取 $k = 0.2/\text{天}$
- W : 年总废水排放量
- F : 年总流量
- R : 年废水总排放量与年总流量之比
- P : 某类水质对应的河段长度占总河长的百分比
- x : 年处理废水量

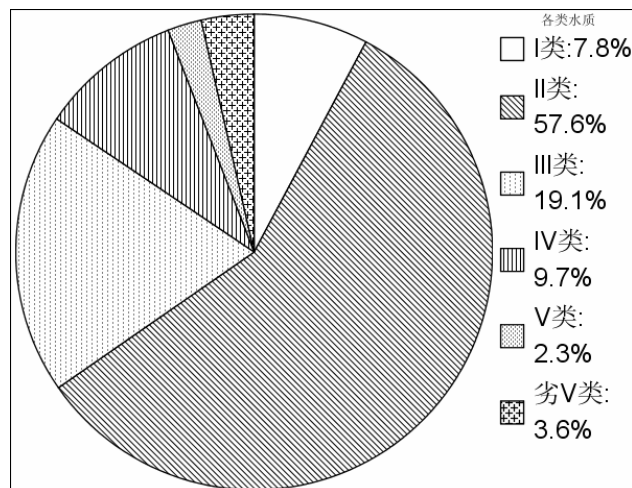
5.模型假设

1. 假设在干流相邻观察站点之间, 排污点沿岸均匀分布, 且每个排污点排出的各污染物的浓度均相同;
2. 不考虑污染物在江水中的弥散作用, 认为污染物排入江水后立即与江水混合均匀。

6.水质综合评价, 地区污染状况: 对问题 1 的回答

6.1 总体概览

对 2003.6.~2005.9.的长江流域的 17 个观测站进行 476 次水质观测数据综合评价, 各类水质所占百分比如下图:



如上图所示，I类水质相当少，仅占 7.8%，超过半数的测验结果属于第 II 类水质，其次有 19.1% 属于第 III 水质，不可饮用的 IV，V 和劣 V 类水质共占到 15.6%

采用加权平均法，将长江干流上的各个观察站的某物质浓度以该站水流量为权重求取平均值，可得到长江干流上该物质的平均浓度。我们以此考察水质分类所用的 3 项物质在长江干流中的平均含量。由于题目仅给出 13 个月的水流量数据，我们选取 2004.5~2005.4 数据，对 7 个观察站水流量以及 3 项物质在观察站处的浓度进行插值计算，再加权平均，得到最近的一年内长江干流上溶解氧、高锰酸盐指数、氨氮的平均浓度，其值分别为 8.26mg/L, 2.62mg/L, 0.25mg/L。可见，长江的水质目前在总体上可划归第 II 类。

6.2 单项目指标评价

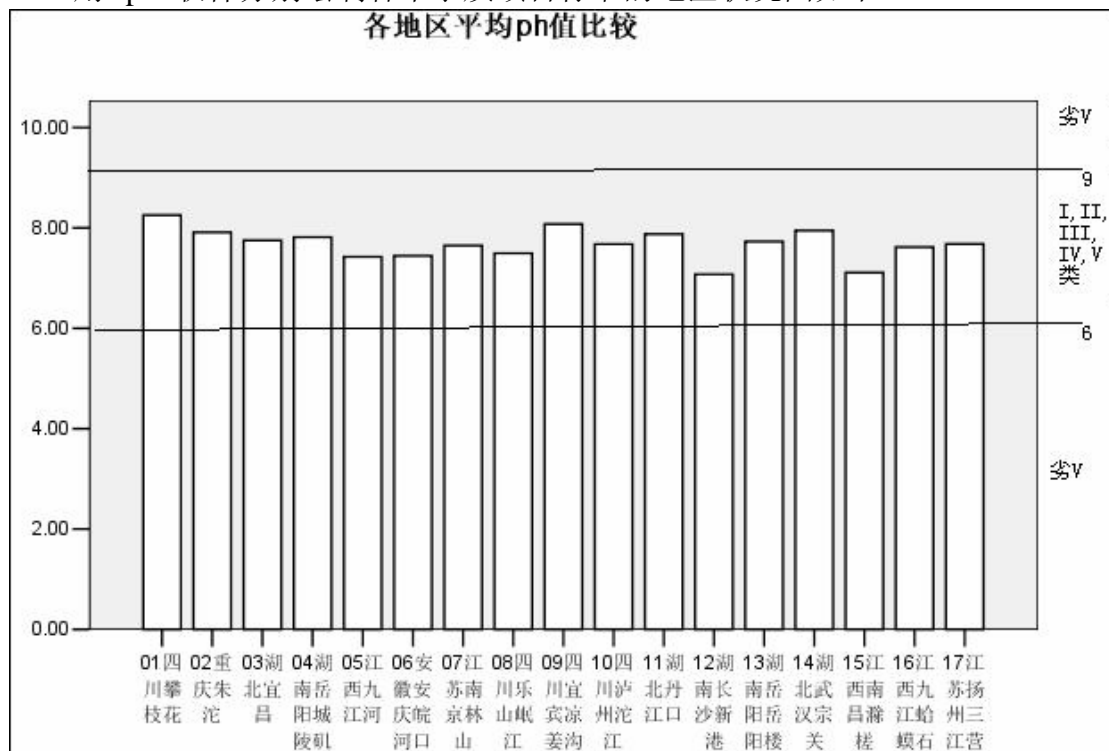
在 476 次测量中，分别以四个项目的标准限值划分水质等级，得出这四种划分方式下各水质等级采样点所占总采样点比例数如下

水质类别	PH 值	溶氧量 (DO)	高锰酸盐指数	氨氮 (NH ₃ -N)
I	100%	56.51%	28.99%	27.52%
II		30.46%	53.78%	50.21%
III		7.14%	13.03%	11.76%
IV		5.46%	4.20%	4.83%
V		0.21%	0	2.10%
劣 V	0	0.21%	0	3.57%

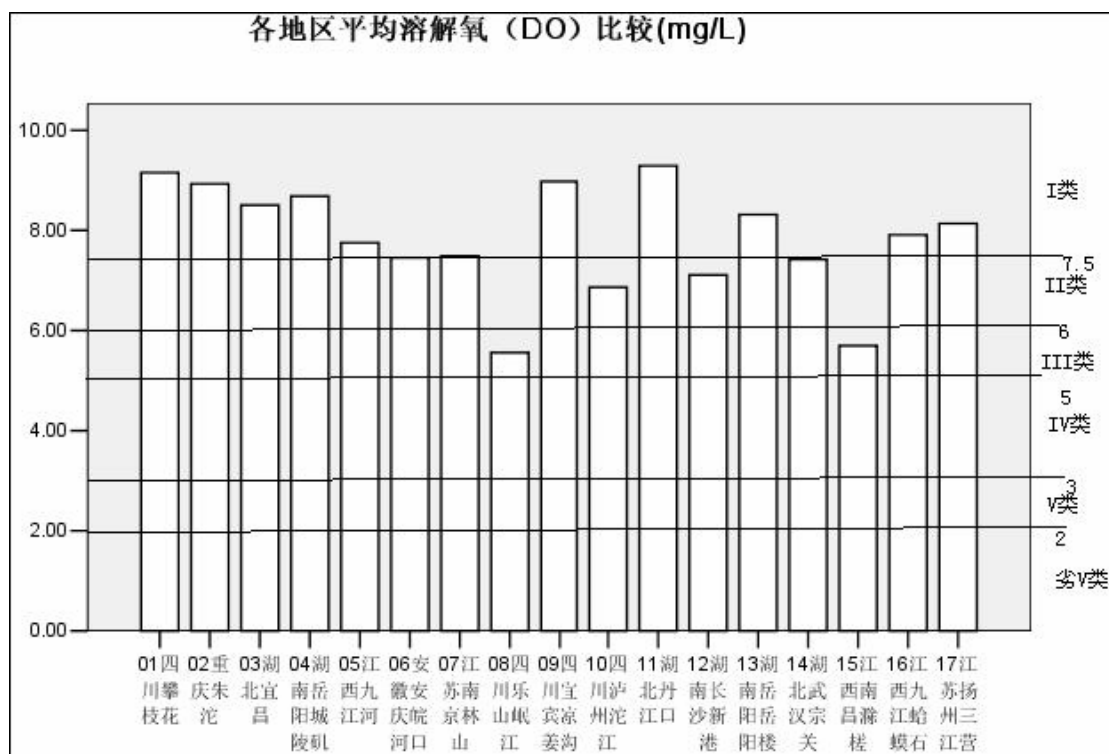
可见 PH 值无超标现象，另外三种指标的第 I，II 类水质均占到 80% 左右，对不可饮用水 (IV, V, 劣 V 类)，其比例分别为 5.82%，4.2% 和 10.5%，可见水域中氨氮 (NH₃-N) 的污染相对严重，高锰酸钾指数 (CODMn) 无低于第 IV 类标准的采样值，溶氧量 (DO) 也极少低于第 IV 类标准。

6.3 地区单项目对比评价

用 spss 软件分别绘制各个水质项目标准的地区状况图如下：

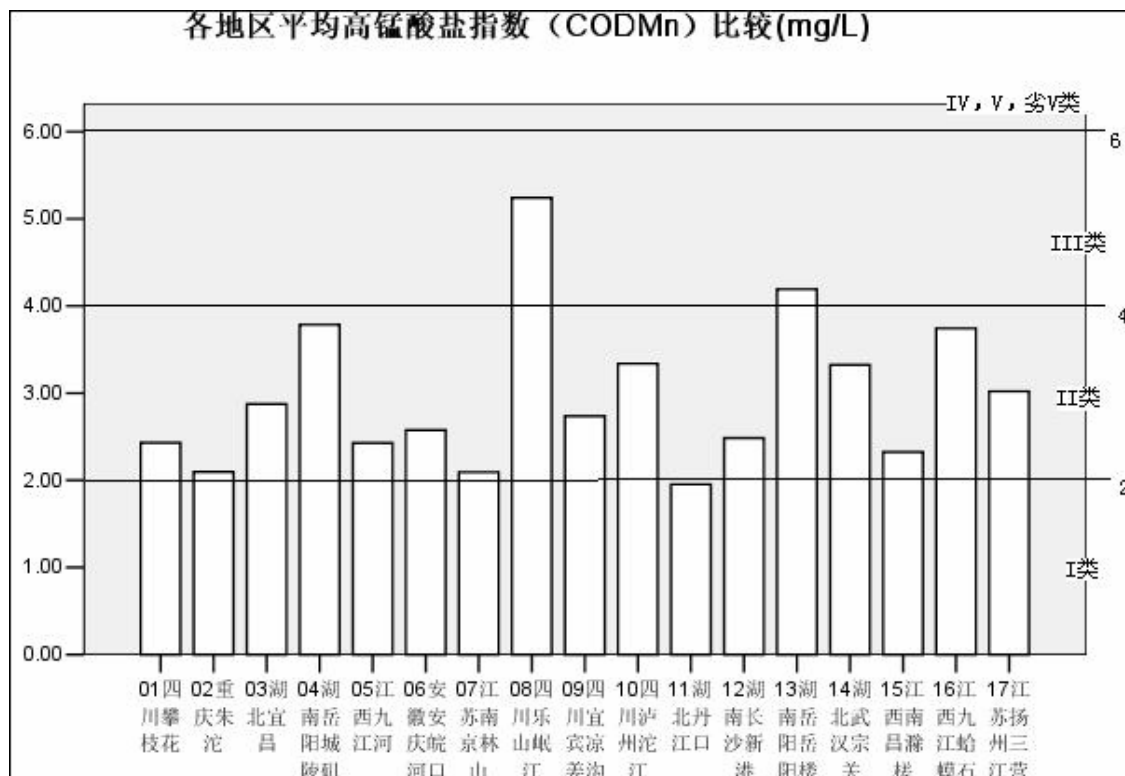


各个地区的 PH 值虽有不同，但都在 6-9 的范围内，并且距离超限还有较大空间；

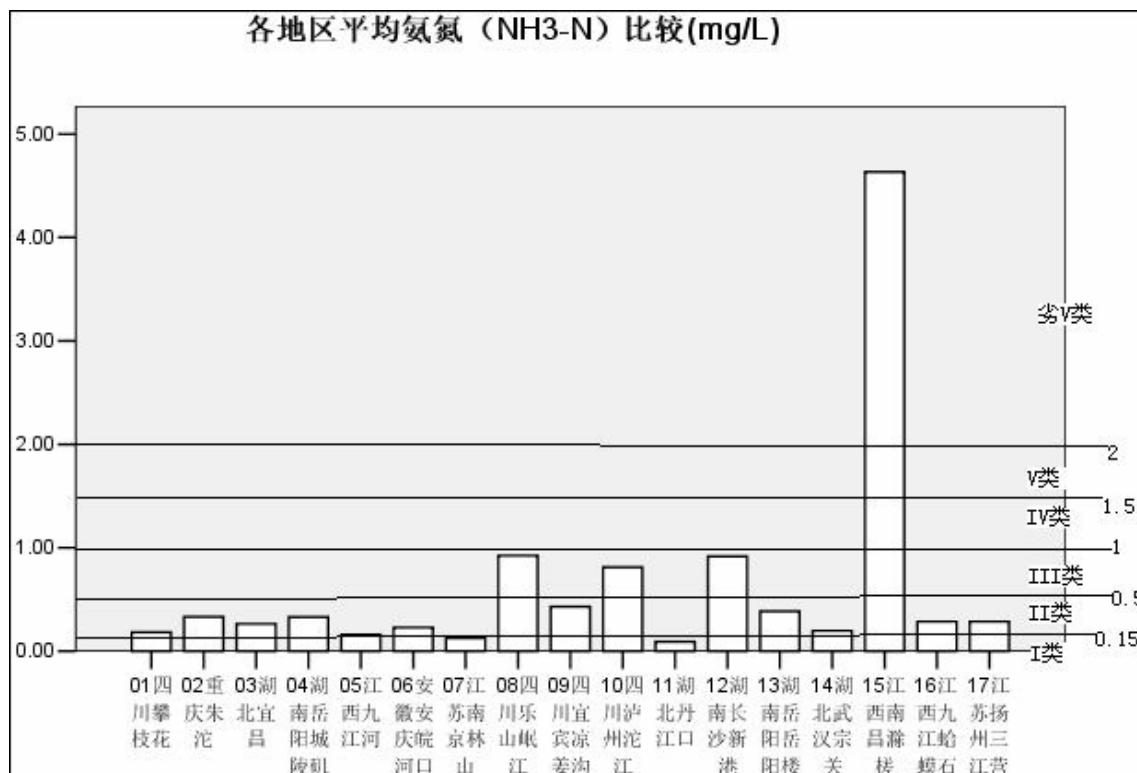


所有地区的平均溶解氧 (DO) 都在第 III 类水质标准以上，即平均为可饮用水，并且除四川乐山岷江大桥和江西南昌滁槎外其它均为第 I, II 类水质；而且可看出干流的

七个观测点的水质要明显优于支流平均状况。



对于平均高锰酸盐指数 (CODMn)，除四川乐山岷江大桥和湖南岳阳岳阳楼为 III 类水质外，其它均为第 I，II 类水质；然而水质最好的地区的平均高锰酸盐指数都接近于第 I 和 II 类的分界线，说明各地区虽污染程度不严重，但普遍被污染。



显而易见，江西南昌滁槎的平均氨氮（NH₃-N）远远高于其它地区，并远远超出劣Ⅴ类水质标准，污染相当严重。属于第Ⅲ类水质的湖南长沙新港、四川泸州沱江二桥和四川乐山岷江大桥也接近Ⅳ类水质标准，其它地区水质相对好，尤其在七个干流区域，平均水质都在Ⅰ，Ⅱ类别。支流的某些区域相对容易受到氨氮的污染。

各地区各项观测指标的平均水平基本都符合饮用水要求，即属于Ⅰ，Ⅱ，Ⅲ类水平，但并不保证所有时间，这些地区的水质都适合饮用，因为各项观测指标的平均水平反映的是各地区不同污染物的大致状况。考虑到各地区水质随时间变化有着很大的不均衡性，为了更严格的确定各地区水质类别，我们采用单指标最差原则，即认为各个地区水质污染状况由其水质最差指标的级别决定。

6.4 综合评价

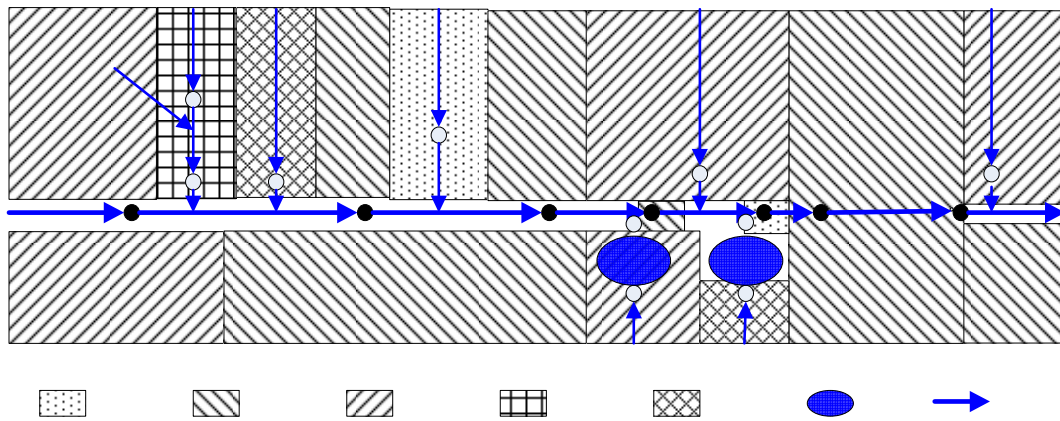
单指标最差原则：

对于地区的每次观测，取 4 个指标下水质最差的类别作为观测结果，取其 28 次的观测结果中水质最差者作为其水质类别。单指标最差原则是对水质的一个严格的限制，该原则下的各地区任何时期的水质水平都严格优于其水质类别要求。即该原则下属于饮用水标准Ⅰ，Ⅱ，Ⅲ类的地区，在任何时间的水质都是饮用安全的。

根据在每个地区进行的 28 次，共计 476 次观测结果，我们得到下表：（单位：次）

	Ⅰ类	Ⅱ类	Ⅲ类	Ⅳ类	Ⅴ类	劣Ⅴ类	水质类别
四川攀枝花	8	15	3	2	0	0	Ⅳ
重庆朱沱	0	20	8	0	0	0	Ⅲ
湖北宜昌南津关	0	25	3	0	0	0	Ⅲ
湖南岳阳城陵矶	0	18	10	0	0	0	Ⅲ
江西九江河西水厂	2	26	0	0	0	0	Ⅱ
安徽安庆皖河口	1	26	1	0	0	0	Ⅲ
江苏南京林山	2	25	1	0	0	0	Ⅲ
四川乐山岷江大桥	0	4	9	9	6	0	Ⅴ
四川宜宾凉姜沟	1	19	4	1	3	0	Ⅴ
四川泸州沱江二桥	3	13	4	4	1	3	劣Ⅴ
湖北丹江口胡家岭	20	8	0	0	0	0	Ⅱ
湖南长沙新港	0	3	14	11	0	0	Ⅳ
湖南岳阳岳阳楼	0	15	10	3	0	0	Ⅳ
湖北武汉宗关	0	18	9	1	0	0	Ⅳ
江西南昌滁槎	0	0	3	10	1	14	劣Ⅴ
江西九江蛤蟆石	0	19	5	4	0	0	Ⅳ
江苏扬州三江营	0	20	7	1	0	0	Ⅳ

在长江一维概化图上，依照上述分类结果得到长江水质的大略分布图：



由上图可看出，长江流域水大部分属于第 III 类和第 IV 类。干流水质优于支流水质，四川岷江，沱江，江西赣江流域污染严重，而南水北调的引水区丹江口水库水质较好。此外，水流经洞庭湖，鄱阳湖后水质均变好。

隶属度评价

在上述单指标最差原则评价中，一个地区水质的综合评价由本河段影响最大的污染因素的指标来衡量，则最差因素相同而其他污染因素水平有差别的两个河段，在此综合评价中的等级是相同的。为了得到河水中各个污染因素对水质的综合影响，我们引入模糊综合评价法对各断面的水质情况做出综合评价。首先对影响水质的各个污染因素进行单独评价，用模糊数学的概念，计算出每个地区的污染因素的平均浓度对于各个水质类别的隶属度。然后考虑不同因素在总体中的地位，配以适当的权重，经过运算得出总的评价评价结果。起计算步骤如下：

Step1: 建立因子集：根据题目所给的影响水质的因素，建立因子集为类 III 类 III 类 IV 类
 $u = \{DO, CODMn, NH_3-H, PH\}$

Step2: 对于每一个因子集里的元素，计算每个地区的相应因素对于各个等级的隶属度。以氨氮 NH_3-H 为例，以 NH_3-H 的浓度作为论域 $U = \{0, +\infty\}$ ，则“I 类水质”、“二类水质”等都是 U 上的模糊子集。每一个确定的 NH_3-H 的浓度 ρ 对于模糊子集“第 i 类水质”的隶属度为 $\mu_i(\rho)$ ，其中函数 $\mu = \mu_i(x)$ 为浓度 x 对于模糊子集的隶属度函数。 NH_3-H 的浓度区间分界线为题目中所给出的《地表水环境质量标准》的六个限值，即 $L_{NH_3} = \{0.15, 0.5, 1.0, 1.5, 2.0, +\infty\}$ ，则定义隶属度函数为：

$$\mu_1(x) = \begin{cases} 1 & x < L_{NH_3}(1) \\ \frac{L_{NH_3}(2) - x}{L_{NH_3}(2) - L_{NH_3}(1)} & L_{NH_3}(1) \leq x < L_{NH_3}(2) \\ 0 & else \end{cases}$$

...

$$\mu_i(x) = \begin{cases} 1 - \mu_i(x) & L_{NH_3}(i-1) \leq x < L_{NH_3}(i) \\ \frac{L_{NH_3}(i+1) - x}{L_{NH_3}(i+1) - L_{NH_3}(i)} & L_{NH_3}(i) \leq x < L_{NH_3}(i+1) \\ 0 & else \end{cases}$$

...

$$\mu_6(x) = \begin{cases} 1 - \mu_5(x) & L_{NH_3}(5) \leq x \\ 0 & else \end{cases}$$

由此可得出此地区的 NH_3-H 的浓度关于 6 个水质等级的隶属度向量。类此可得出其他污染因素的隶属度向量

Step3: 设四个因素对总体水质影响的权值分别分 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ ，将同一个地区四个污染因素的隶属度向量以 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 为权值求加权平均，则可得到 17 个地区的整体水质对于六个水质类别的隶属度。

在此以 $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1/4$ ，求的 17 个地区的整体水质的隶属度向量如下

	I 类	II 类	III 类	IV 类	V 类	劣 V 类
四川攀枝花	0.90	0.10	0	0	0	0
重庆朱沱	0.81	0.19	0	0	0	0
湖北宜昌南津关	0.75	0.25	0	0	0	0
湖南岳阳城陵矶	0.53	0.47	0	0	0	0
江西九江河西水厂	0.92	0.08	0	0	0	0
安徽安庆皖河口	0.82	0.18	0	0	0	0
江苏南京林山	0.98	0.02	0	0	0	0
四川乐山岷江大桥	0	0.36	0.64	0	0	0
四川宜宾凉姜沟	0.61	0.39	0	0	0	0
四川泸州沱江二桥	0.30	0.49	0.21	0	0	0
湖北丹江口胡家岭	1.00	0	0	0	0	0
湖南长沙新港	0.50	0.22	0.28	0	0	0
湖南岳阳岳阳楼	0.44	0.53	0.03	0	0	0
湖北武汉宗关	0.72	0.28	0	0	0	0
江西南昌滁槎	0.28	0.29	0.10	0	0	0.33
江西九江蛤蟆石	0.58	0.42	0	0	0	0
江苏扬州三江营	0.70	0.30	0	0	0	0

从上表看出，整体水质隶属度的提出，可以避免传统的仅用一个确定性指标衡量而导致对水质总体污染程度性缺乏认识的现象，使评价更为全面和客观。此表即可作为个地区污染水平的一个总的评价，也可作为采取污水处理工作优先级别的确定。

但这种基于模糊数学的评价标准也存在其固有的缺陷。首先，由于我们对各个不同污染因素的影响做加权平均，如果在某一河段，某一种污染物的浓度显著的高，而其它

污染物水平相对较低,会导致两种污染物浓度对应的污染程度的模糊子集没有交集甚至相差很远;此时,加权后整体水质的隶属度集合中会出现断层。如江西南昌滁槎的数据,由于此地的 NH_3-H 水平显著偏高,导致其最终的整体水质的隶属度集合出现断层,显然,这是不合理的。

同时当某些水域个别污染物质显著偏高而且它污染物浓度相对较低时,由于对于每一个地区此标准只给出一个综合的水质的加权均值,在一定程度上会淡化此显著污染物的影响而将此流域划为水质相对较高的等级。而据我们所知,任何主要污染物浓度高于一定阈值时,河流的水是不适合饮用的,但上述的综合评价并没有很好的反映这一点。

因此,在涉及饮用水评定标准时,单指标最差原则比隶属度函数法标准更具合理性。

7.污染源分析: 对问题 2 的回答

本部分我们考察一年来长江流域高锰酸钾指数和氨氮的主要污染来源,以七个干流观察站为分界,将长江划分为六个区段。而对在每个区段内汇入长江的支流所经流域,我们将其划分至本区段范围内,即每个区段的污染物排放来源包括本区段干流上排放源和本区段内所汇入支流上的排放源。

在水文学中,水环境的污染源可分为点源和非点源,点源指污染物集中排入水体的污染源,非点源指排放方式不确定的污染源。非点源按照排放方式分为以体积方式排放,以线源方式排放和以面积源排放三种^[1]。在本题中,根据模型假设,沿岸均匀分布排污点,且污染物一旦进入江水立刻混合均匀,即不考虑污染物在江水中的弥散作用,故我们可将干流沿岸均匀分布的诸个点源等同视为以线源方式排放的非点源,以流量不可忽略的线源污染模型求解,支流可视作污染源处理。最终求出各个区段污染物的单位时间排放总量。

7.1 线源污染模型

分析河流污染物在流过 Δx 长度后污染物质量的变化。令 A_c 为河流的横断面积, v 为河流流速,不考虑污染物的弥散作用,由质量守恒方程可得:

$$C_d \cdot \Delta Q = (Q + \Delta Q)C(x + \Delta x, t + \frac{\Delta x}{v}) - QC(x, t) \xrightarrow{\Delta x \rightarrow 0} (Q + \Delta Q)C(x + \Delta x, t) - QC(x, t)$$

$$\text{则} \quad \frac{d(QC)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{(Q + \Delta Q)C(x + \Delta x, t) - QC(x, t)}{\Delta x} = C_d \frac{dQ}{dx}$$

再考察污染物浓度随时间的变化关系。在某一点,有 $\frac{dC}{dt} = kC$, 表示污染物的降解速率与浓度成正比, k 即降解系数。而污染物浓度既随时间变化,也随空间变化。为得到浓度变化的时空表达式,将以上两公式的导数转为偏微分,作统一量纲处理后相加,得到:

$$\frac{\partial(A_c C)}{\partial t} + \frac{\partial(QC)}{\partial x} = \frac{\partial Q}{\partial x} C_d - k A_c C$$

对线污染源,有 $\frac{\partial Q}{\partial x}$ 为常数,不妨设 $Q = Q_0 + qx$ 。在本题中,我们以一个月为时间

单位研究污染状况，在一个时间单位内，认为河流的污染物浓度不发生变化。则有：

$$\frac{\partial((Q_0 + qx)C)}{\partial x} = qC_d - kA_c C$$

设初始条件为 $v_0 = u, A_{c,0} = A_0$ ，为得到此关于 C 的微分方程的解析解，考虑两种简化情况：

(1) 流速不变，水流量仅影响河流横截面积， $A_c = Q/u$

微分方程为：

$$(Q_0 + qx) \frac{dC}{dx} + (q + k \frac{Q_0 + qx}{u})C - qC_d = 0$$

解之得：

$$C = C_0 \exp(-\frac{kx}{u}) + C_d \frac{q}{A_0 k} (1 - \exp(-\frac{kx}{u})) \frac{u}{u + qx/A_0}$$

对相距为 L ，污染物浓度分别为 C_1, C_2 的两观察站间的地区，将 $x = L, C_0 = C_1, C = C_2$ 代入，得到所接纳水流平均污染物浓度：

$$C_d = -\frac{k[uC_1 - (u + qL/A_0)C_2 \exp(kL/u)]}{(\exp(kL/u) - 1)uv}$$

(2) 河流横截面积不变，水流量仅影响流速， $v_0 = Q/A_0$

微分方程为：

$$(Q_0 + qx) \frac{dC}{dx} + (q + kA_c)C - qC_d = 0$$

解之得：

$$C = C_0 (\frac{u}{u + ax})^{(k+a)/a} + C_d (\frac{a}{k+a}) (1 - (\frac{u}{u + ax})^{(k+a)/a})$$

其中， $a = q/A_0$

对相距为 L ，污染物浓度分别为 C_1, C_2 的两观察站间的地区，所接纳水流平均污染物浓度为：

$$C_d = \frac{(k+a)[(u+ax)C_2 - u(\frac{u}{u+ax})^{k/a} C_0]}{(u+ax - u(\frac{u}{u+ax})^{k/a})}$$

对应每个月的水流量与污染物浓度数值，可得到第 j 个月第 i 个地区的 $C_d(i, j)$ 。由浓度与质量间换算关系，即可计算出两观察站之间地区平均单位时间排放污染物量：

$$M_i = \frac{1}{12} \sum_{j=1}^{12} C_d(i, j) \Delta Q(i, j)$$

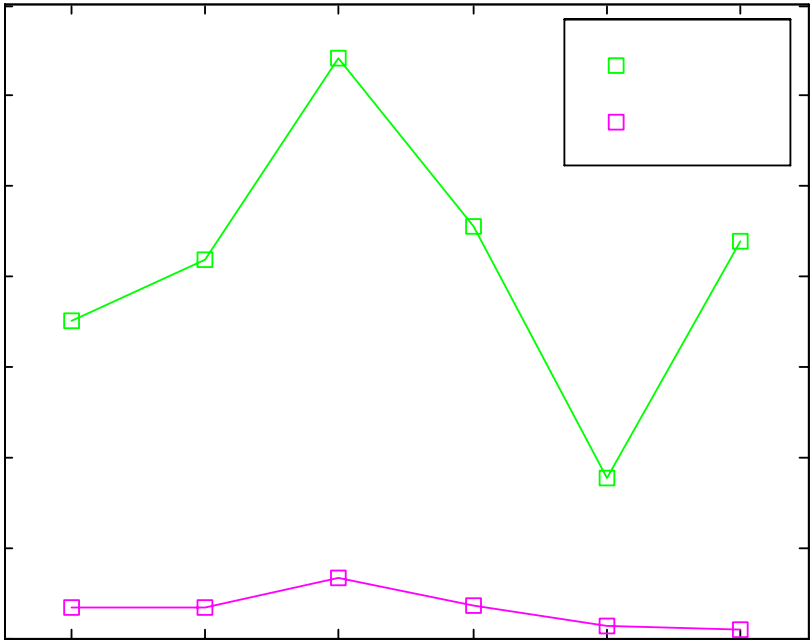
以上推导的是增加流量的非点污染源模型，我们以流量线性增加求解，但同样适用于其他非点源。

在实际中，河流的流速与横截面积均随距离变化，在此情况下无法给出上述微分方

程的确定解析式。为此，可根据河流的状况选取符合实际的计算方法。在本题中，由于长江干流水流速变化缓慢(相邻两观察站流速差值在所给记录资料中平均为 $0.54m/s$ ，为平均流速的 29.6%，尤其在宜昌以下中下游四个地区，流速差值相当小，其平均值仅为 $0.25m/s$ ，为平均流速的 17.2%)，故我们采用流速不变方法，给出各地区污水排放量的解：

各地区单位时间平均污染物排放量 (流速不变法)

项目 \ 地区	地区一	地区二	地区三	地区四	地区五	地区六
	攀枝花-重庆	重庆-宜昌	宜昌-岳阳	岳阳-九江	九江-安庆	安庆-南京
高锰酸钾	3501.5	4180.1	6416.1	4558.4	1767.8	4385.5
排名	5	4	1	2	6	3
氨氮	340.81	340.59	656.82	353.21	135.83	89.51
排名	3	4	1	2	5	6



在宜昌至岳阳之间，由高锰酸盐指数计算的污染物排放量与氨氮排放量均为所有地区中的最大值，分别占有所有地区的总和的 25.9%和 34.3%，岳阳到九江之间两项污染物排放量均位于第二位，分别占有总和的 18.3%和 18.4%，这两个地区即长江干流一年来高锰酸盐指数和氨氮的主要污染源。

(t/day)

8.预测模型的建立和求解：对问题 3 的回答

8.1 对六种类别的聚类分析：

国家标准将水质分为 I, II, III, IV, V 和劣 V 六类，对于预测未来十年水质标准的六个类别的百分比变化，没有特别大的参考意义并且难以精准预测，于是对现有的六个水质类别按照各个项目观测临界值进行 Q 型聚类分析，

注：对于劣 V 类水的 ∞ 标识，使用相对大的数代替，对 PH 值，前五类与劣 V 使用采用不同数值

用欧氏距离平方 (Squared Euclidean Distance) 作为 6 个水质分类之间亲疏程度衡量指标，采用类间平均链锁法 (Between-groups Linkage) 表示原类别与新类，新类与新类之间的亲疏程度

$$\text{欧氏距离平方} = \sum_{i=1}^k (x_i - y_i)^2,$$

其中， k 表示原类别对应的变量数，这里为 DO, CODMn, NH3-N 和 PH 值， $k = 4$

x_i 表示原类别 x 在变量 i 上的取值， y_i 表示原类别 y 在变量 i 上的取值，这里分别一个项目指标下两个原类别对应的临界值

类间平均链锁法：将两个新类之间的距离表示成两个新类内所有原类别间的平局距离

在此基础上得到垂直冰柱图如下左图

Vertical Icicle											
Number of clusters	Case										
	劣V类		V类		Ⅳ类		Ⅲ类		Ⅱ类		I类
1	X	X	X	X	X	X	X	X	X	X	X
2	X		X	X	X	X	X	X	X	X	X
3	X		X	X	X		X	X	X	X	X
4	X		X		X		X	X	X	X	X
5	X		X		X		X	X	X		X

CASE_LBL	Cluster	Distance
I类	1	2.437
II类	1	.174
III类	1	2.359
IV类	3	2.562
V类	3	2.562
劣V类	2	.000

左图表格中行标为分类数 (1-5)，代表希望分类的个数，某行中连续的×号表示对应连续的若干个与原类别聚合为一个新类。

右图中更为直观的显示了当聚合成三个新类时各个原有水质类别的归属，第三栏代表原有类别与其新类别的中心点距离。

由此我们为了更加合理有效的预测十年的水质变化将水类别重新分为三类：

I, II, III 类为可饮用水，称为一等水；

IV 和 V 为轻度污染水，称为二等水；

劣 V 是重度污染水，称为三等水。

对于 2005~2014 年的变化规律的预测，我们准备采用两种数理统计方法进行预测，

分析水文要素自身的统计规律或要素与有关因子之间的统计关系，然后应用这些规律或关系设计预测方法。

<1>时间序列分析方法:该方法从水文要素的自身统计规律入手，把水文要素的历史变化作为时间序列处理，利用要素的历史演变规律作外推预报。

<2>多元分析方法:该方法把前期各影响因子作为随机变量，通过统计分析找出它们与预报要素之间的统计关系，利用前期因子预报未来水文要素的取值。

8.2 时间序列分析预测：指数平滑法

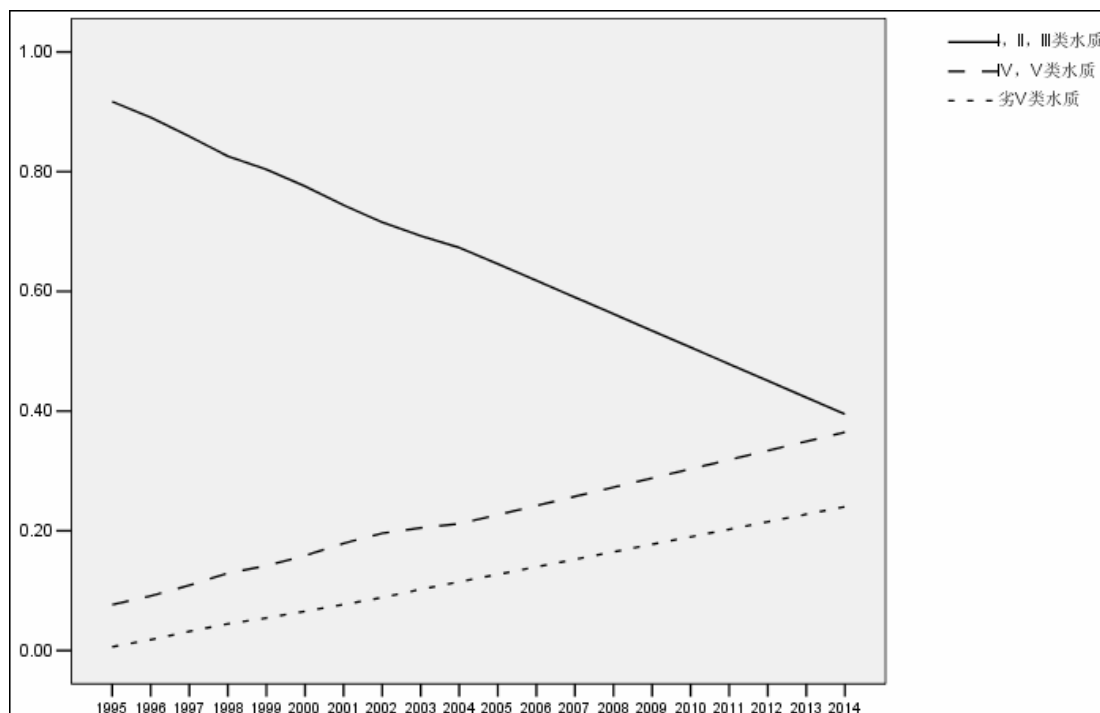
观察 1, 2, 3 类水质 1995-2004 全年所占河长百分比的变化，认为其具有一定的平稳性和趋势性，所以将来各个水质所占百分比也应具有相同的平稳性和趋势性，

平稳性体现在未来某年的某类水质所占百分比 $P(j)$ 需要所有先前年份水质百分比 $P(i), i < j$ 加权决定，趋势性体现在 $P(j)$ 与其临近的若干个先前年份的水质关系较大，也就是最近的年份权系数要大

基于以上几点考虑，我们用时间序列的指数平滑法进行加权

$$P(t+1) = \alpha \cdot [\sum_{k=0}^t (1-\alpha)^k P(t-k)]$$

其中 α 为平滑常数， α 可在 $[0,1]$ 之间取任何值。各个权系数随时间变化呈现一条衰减的指数函数曲线，即随时间向过去推移，各个时间历史值对预测值的影响按照指数规律递减。当平滑常数接近 1 时，各期历史数据的作用将迅速衰减，近期数据对预测值的作用最大。使用 spss 软件进行时间序列指数平滑预测（spss 根据拟合误差大小自动搜索最合适的 α 值）得到 1, 2, 3 等水质占河长百分比的十年预测如下



显而易见，随着时间推移，I，II，III 类水质百分比逐渐下降，而 IV，V 类和劣 V 类水质百分比逐年上升，可以看出到 2014 年，长江流域的可饮用水仅仅不到 50%，生态恶化形式相当严峻。

上述预测可以看出，指数平滑法体现了百分比变化的平稳性和趋势性，但实际上，各个水质百分比的变化还可能受制于周期性因素。可是对于 1995-2004 仅十年的数据量无法使用指数平滑法拟合并预测，使其同时具有平稳性，趋势性和周期性。这也就是 2005 年之后的预测值成直线性的原因。

对于周期性因素影响百分比变化的规律，我们需要挖掘各水质百分比的形成的影响或相关因子，对各水质百分比进行机理分析。

8.3 时间序列分析预测：神经网络

Kolmogorov 定理证明：三层神经网络的输出可逼近任意连续实值函数。而任一时间序列都可看成是由非线性机制确定的输入输出系统^[2]。因此，将神经网络应用于时间序列具有较强的可行性。

在本题中，我们采用基于反向传播算法的多层感知器网络，进行长江水质变化时间序列的预测。在第一步在数据预处理中，由于在预测问题中，将来值可能超过当前已记录值范围，不适宜使用最大最小归一化，我们采用了标准化，即将输入向量 \mathbf{x} 标准化为

$$\frac{\mathbf{x} - \bar{\mathbf{x}}}{mse(\mathbf{x})}, \text{ 其中 } mse(\mathbf{x}) = \frac{1}{N}[(\mathbf{x} - \bar{\mathbf{x}})^2]^{1/2} \text{ 为 } \mathbf{x} \text{ 的均方差。}$$

网络输入设计为六类水河长占总河长百分比在上一年份的值，输出为本年份的这六个值。以全流域水文年 1995-2004 年数据作为训练样本，选取双曲正切 S 型传递函数进行学习。由于训练样本较少，采用交叉验证法，其规则为将输入样本平均分为 n 组，将其中的 $n-1$ 组样本点作为训练集合，剩余 1 组作为验证集合，并作 n 轮实验，在每一轮实验中以轮换交替原则选取训练和验证集合，最后对各轮实验的结果取算术平均值。交叉验证法能有效提高网络的学习性能^[3]。

将网络输出值反标准化得到预测值(单位：%)：

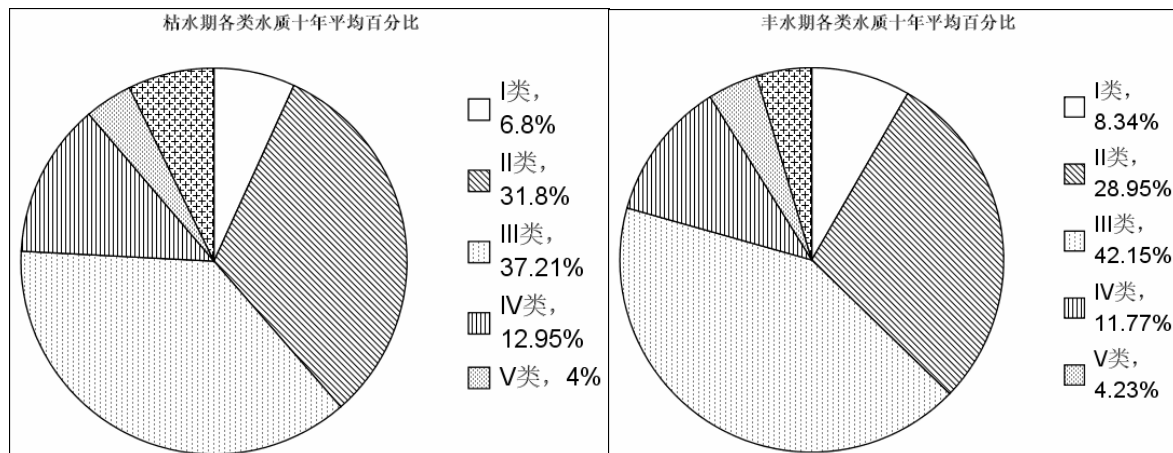
	I 类	II 类	III 类	IV 类	V 类	劣 V 类	一等	二等	三等
2005 年	1.02	31.26	36.08	16.54	3.919	10.92	68.36	20.459	10.92
2006 年	1.26	26.58	29.1	16.86	10.23	15.47	56.94	27.09	15.47
2007 年	1.06	22.41	25.71	20.58	14.8	16.64	49.18	35.38	16.64

2005-2007 年的预测值与指数平滑法预测值基本一致，但此后年份的预测值相对很不稳定，没有取得预想的预测效果。由神经元对输入加权求和的原理知，每年的预测值受输入的前一年份的值决定，这将造成误差累积，使输出误差逐渐增大。

8.4 多元分析预测：线性预测模型

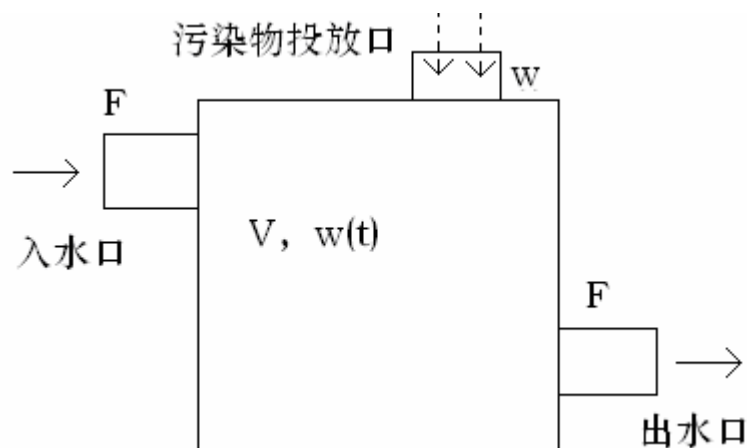
为了得到水质变化客观的内在的规律，我们要寻求不同等级水质的河段所占的比例与什么因素相关。首先，不同等级水质的河段比例是对长江水质污染程度的一个总体刻画，

而污染程度是与废水排放总量 W 密切相关的。其次，基于对丰水期和枯水期水流量和水质的统计，我们发现丰水期的水质普遍比枯水期的水质优良（如下两图所示）。于是，我们选取年废水排放量 W 与年总流量 F 作为研究对象,探讨此两个因素对于水质的影响.



8.4.1 年废水排放量 W 与年总流量 F 对水质的函数关系的猜想

为了讨论年废水排放量 W 与年总流量 F 对水质的函数关系,我们构造了一个简单的恒定流量水箱模型来模拟长江的水流状况,如下图:



我们用上图的水箱来抽象整个长江，水箱的出水口与入水口分别是对长江所有入水口与出水口集合的简单抽象，类似，水箱的污染物投放口代表长江排污口的总和。我们三个假设：

- 1.在讨论的时间范围内，水箱内水的体积不随时间变化
- 2.在讨论的时间范围内，水箱的入水速率、污染物投放速率不随时间变化
- 3.只有一种污染物

具体的，我们用水箱的贮水量 V 来表示长江的总水量(m^3)；用 f 表示单位时间注入水箱的水量(m^3/s)。由假设 2， f 可以理解为长江在一个相对长的时间内（一年）的平均流量。同时，在总水量 V 不变的假设下，可认为流出水箱的水的流量也为 f ；定义 w

为单位时间内投放污染物的质量 (kg)，即长江在一个相对长的时间的平均污染物投放量的一个刻画。相应的，定义水箱中时刻 t 的污染物的总量为 $w(t)$ ，污染物浓度为 $\rho(t) (mg/m^3)$ 。其中 $\rho(t)$ 即为水箱中污染程度的刻画，也可认为是长江水质的一个度量。

讨论在较短的一段时间 Δt 内，水中的污染物总量 $w(t)$ 的变化应满足：

$$w(t + \Delta t) - w(t) = V \cdot (\rho(t + \Delta t) - \rho(t)) = \Delta t \cdot w - \Delta t \cdot f \cdot \rho(t)$$

整理得：

$$\frac{V \cdot (\rho(t + \Delta t) - \rho(t))}{\Delta t} = w - f \cdot \rho(t)$$

化为微分方程的形式有：

$$\frac{\partial \rho}{\partial t} = \frac{w}{V} - \frac{f}{V} \cdot \rho$$

由于水箱的入水量 F 起到降低污染物密度的作用，而排放量 W 起到加剧污染状况的作用，且这两个量是恒定的，则应存在一个污染物浓度的平衡点时刻 t' ，使 $\frac{\partial \rho}{\partial t} \big|_{t=t'} = 0$ ，

集水箱达到污染和排污的平衡点使水箱内污染物浓度不再变化，此时时有 $\rho(t') = \frac{w}{f}$ 。可

认为这是长江一个较长时间内污染物平均水平的刻画。在这个意义上，有：

$$\bar{\rho} = \rho(t') = \frac{w}{f} \approx \frac{\bar{w}}{\bar{f}} = \frac{\bar{w} \cdot T}{\bar{f} \cdot T} = \frac{W}{F}$$

其中， \bar{w} 与 \bar{f} 分别为长江的平均排污量和平均流量， T 为一年的时间。这上式的意义在于，如果我们认为稳态时的污染物浓度 $\rho(t')$ 反映了长江一年的平均污染物水平 $\bar{\rho}$ ，同时 f 与 w 也近似地刻画了长江在一个相对长的时间内（一年）的平均流量和平均污染物投放量，则平均污染物浓度 $\bar{\rho}$ 等与一年内中的污水排放量与总流量的比值（经过量纲变化），这个比值我们定义为单位流量的污水排放量 R 。

$$R = \frac{\text{年废水排放量}}{\text{年水流量}} = \frac{W}{F}, \text{ 单位：吨/米}^3$$

由常理，水中的污染物浓度越大，三等水河段的比例越大，一等水河段的比例越低，二等水河段的比例为 100% 检去上两个量的余值。基于以上的分析，我们猜想如下：

1. 一等水河段的比例与单位流量的污水排放量 R 正相关
2. 三等水河段的比例与单位流量的污水排放量 R 负相关

8.4.2 单位流量的污水排放量 R 与各水质等级比例的函数关系确定

为了验证猜想的真实性，我们分别对于 1995~2004 年长江流域 I, II, III 类, IV, V 类和劣 V 类水在总流域中的百分比数值作和 R 作相关性分析得到：

相关性

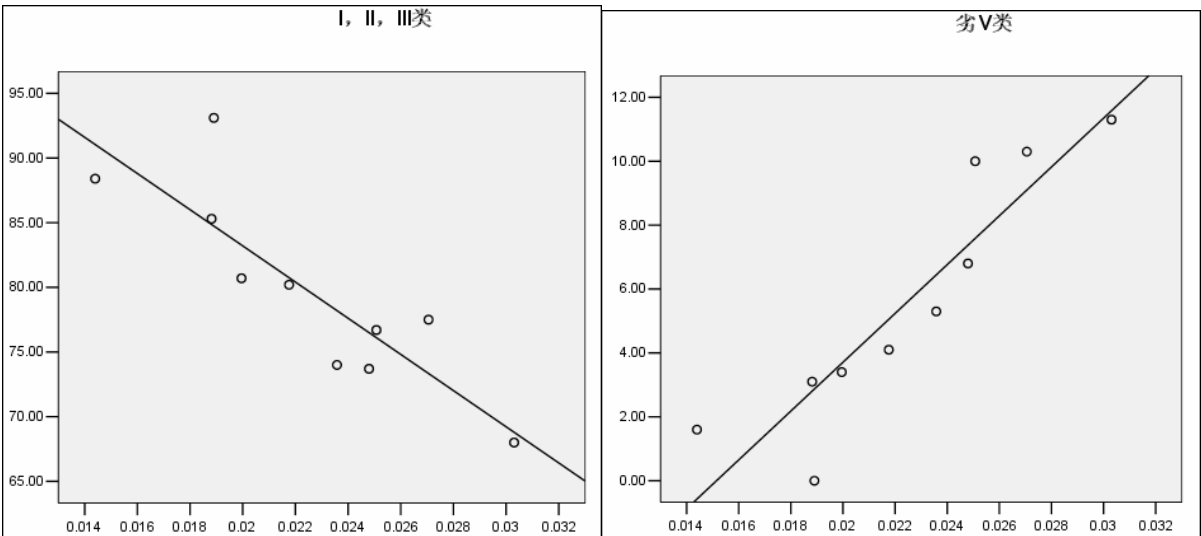
		R	I, II, III类	IV, V类	劣V类
R	Pearson 相关性	1	-.819**	.511	.957**
	显著性（双侧）		.004	.131	.000
	N	10	10	10	10
I, II, III类	Pearson 相关性	-.819**	1	-.893**	-.847**
	显著性（双侧）	.004		.001	.002
	N	10	10	10	10
IV, V类	Pearson 相关性	.511	-.893**	1	.517
	显著性（双侧）	.131	.001		.126
	N	10	10	10	10
劣V类	Pearson 相关性	.957**	-.847**	.517	1
	显著性（双侧）	.000	.002	.126	
	N	10	10	10	10

** . 在 99%置信 水平（双侧）上显著相关。

表中数值的第一项为横行变量与纵行变量的相关系数，第二项为不相关假设的显著性水平。可见在废水总排放量与年总流量之比 R 和一等水质所占的百分比相关系数为 -0.819，显著性水平<0.01，即在 99%的置信水平下认为二者显著负相关，并且相关系数绝对值 $0.819>0.8$ ，视为高度线性相关；

同时，废水总排放量与年总流量之比 R 和二等水质所占的百分比相关系数为 0.957，显著性水平<0.01，即在 99%的置信水平下认为二者显著正相关，并且相关系数绝对值 $0.957>0.8$ ，视为高度线性相关；

二者线性拟合效果如下图所示：



若上述直线反映的是真实的客观的规律，则观测点与直线的误差可以理解为大量随机因素产生的误差；由概率论相关知识可知，这些误差的和应该服从正态分布。即，如

果上述线性关系成立，用最小二乘法拟合直线的残差应服从正态分布。我们对二者残差进行单样本 K-S 检验，得到二者残差分别服从均值为 0，标准差为 3.76（单位：百分比）和均值为 0，标准差位 1.559（单位：百分比）的正态分布，可认为拟合的直线是单位流量污染量 R 与一、二、三等水所占的百分比关系客观规律的高度近似。

由以上分析，我们建立线性预测模型为：

$$\begin{cases} P_1(t) = a_1 R(t) + b_1 \\ P_2(t) = 1 - P_1(t) - P_3(t) \\ P_3(t) = a_3 R(t) + b_3 \\ t = \{ \dots, 1995, 1996, 1997, \dots \} \end{cases}$$

模型说明与分析：

1. $P_1(t)$, $P_2(t)$, $P_3(t)$ 分别表示为 t 年份一等水、二等水、三等水河段占全流域河段的百分比。 $(a_1, b_1), (a_3, b_3)$ 分别表示一等水，三等水线性拟合多项式的系数。

2. 由以上拟合直线可以看出， P_1 与 R 负相关，则系数 a_1 应为负值。其意义可理解为单位流量污染总量 R 减少 1 个单位时，一等水河段长度增加的比例数为 a_1 ，它反映了污染减少或水流量增大对水质净化的贡献率； b_1 为污染物投放量 W 为 0 或水流量极大时一等水质河段的比例，理论上 b_1 应近似等于 100%。

3. 同理分析， P_3 与 R 负相关，则系数 a_3 应为负值。其意义可理解为单位流量污染总量 R 增加 1 个单位时，三等水河段长度增加的比例数为 a_3 ，它反映了污染物投放量增加或水流量减少对水质净化的贡献率；其中当 $P_3 = 0$ 时， $R = -\frac{b_3}{a_3}$ ，这是将三等水比例降为零可以接受的最大单位流量污染。

4. 由于随着 R 值的减少（增大），一等水的比例增大（减少），三等水的比例减少（增大），而二等水的比例为 100% 与上两个值之差，一增一减，变化的效果会弱化；即，二等水河段的比例随 R 值的变化应该不会太明显。

线性拟合的结果为：

$$\begin{cases} P_1 = 1.1118 - 13.98 \cdot R \\ P_2 = 1 - P_1 - P_3 \\ P_3 = 7.64 \cdot R - 0.1157 \\ 0.015 \leq R \leq 0.0795 \end{cases}$$

（ $0.015 \leq R \leq 0.0795$ 的约束是三个比例的非负约束，其中 R 的量纲为吨/米³）可见，拟合结果和理论分析是十分符合的。

预测模型建立后，我们只需得出未来十年每年的污水排放总量和水流量的预测值，就可以运用模型求出十年内三个等级的水质对应河段长度的比例。

8.4.3 年污水排放量 W 的预测

长江污水的主要来源有两类：工业污水与生活污水。工业污水的排放量与长江沿岸的工厂的数量与规模相关，生活污水的排放与长江流域居民数相关；而从经济增产满足人口增长带来需求(产品、服务、就业岗位)增长的角度上来说，生产单位的规模也应与人口正相关。我们从国家统计局得到 1995—2003 年中国人口数据并整理出长江流域省份的常住居民人数^[4]，如下表：

年份	1995	1996	1997	1998	1999	2000	2001	2002	2003
人口(万)	46757	47161	47467	47882	48208	48193	48864	49092	49455

将以上人口数据与污水排放量的数据做相关分析，得到：

人口与污水排放量的相关系数为 0.93，在 99%置信水平下是强相关的，可以假定污水的排放量与长江流域人口数 $popu$ 为线性关系，即认为

$$W = Q(popu) = c_1 \cdot x + c_0$$

其中，参数 c_1 、 c_0 可由最小二乘法确定，我们只需确定未来十年内长江流域的人口数，就可以确定十年内每年污染物的排放总量。

我们利用人口增长的指数模型，即

$$popu(t) = x_0 e^{rt}$$

其中，以 1995 年为 $t=1$ ，将 1995—2003 年的人口数做对数化处理，再以最小二乘法拟合出人口基数 x_0 与增长率 r 的值，得出人口随时间变化的预测函数。

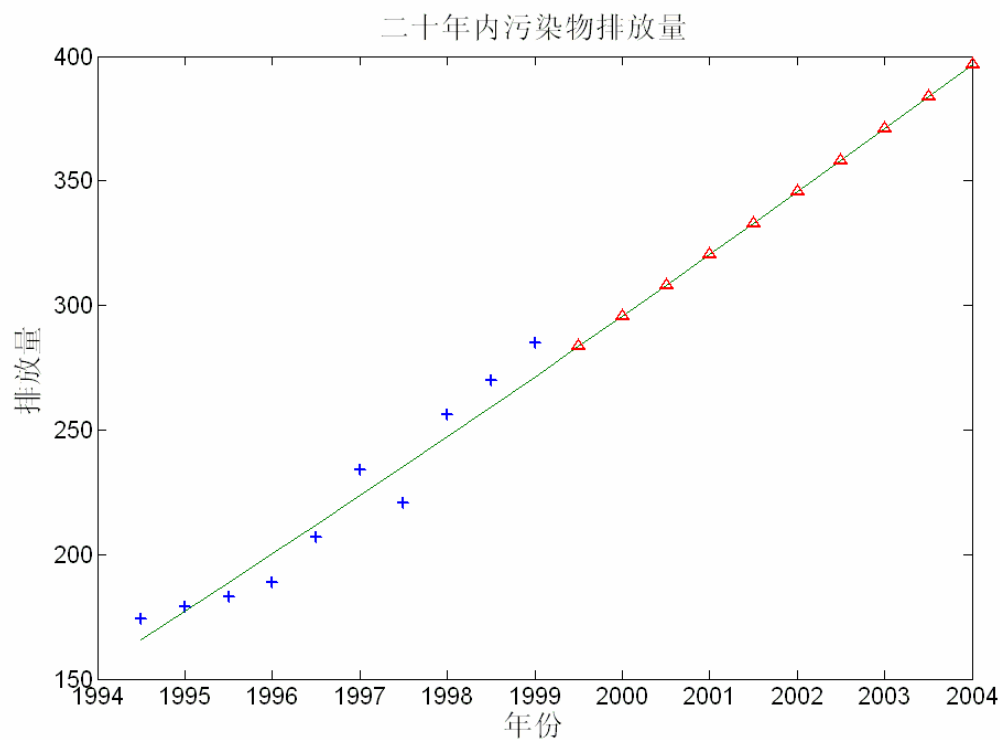
合并上文中的两个公式，可整理出污水排放量随时间变化的预测函数为：

$$W = c \cdot \exp(r \cdot t) + c_0$$

c 、 c_0 与 r 的值分别分 $c=1613$, $c_0=-1458.6$, $r=0.007$ 。其中， r 为人口的增长率，于是也是污染物投放量的增长率； $c+c_0$ 应为为初始污染物投放量的初始值，即 1994 年污染物的投放量。依据以上函数关系确定未来十年的污水排放量 W 的预测值(单位：万吨)：

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
废水量	283.5	295.7	308.1	320.5	332.9	345.6	358.2	371.0	383.4	396.8

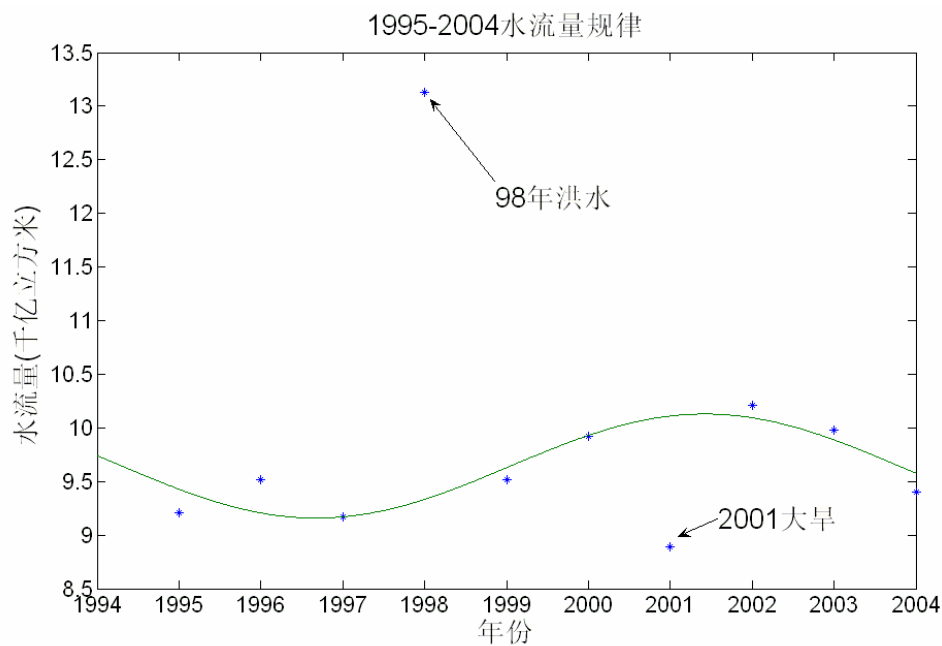
对应曲线为：



图中，星型点为原始数据，带三角的点为预测数据。可见，在十年较短的时间内，可认为污水排放量近似呈线性增长。

8.4.4 水流量 F 的预测

影响水流量变化的因素很多，除了降水量、冰山溶解速率变化等自然因素外，还有人类对水的消耗和水利工程建设等人为因素影响。随机性很强，规律性不明显，预测难度大。因此,我们放弃了对水流量机理分析的尝试,而仅仅拟和正常年份下水流量的变化规律，表征水流量的年度波动性。



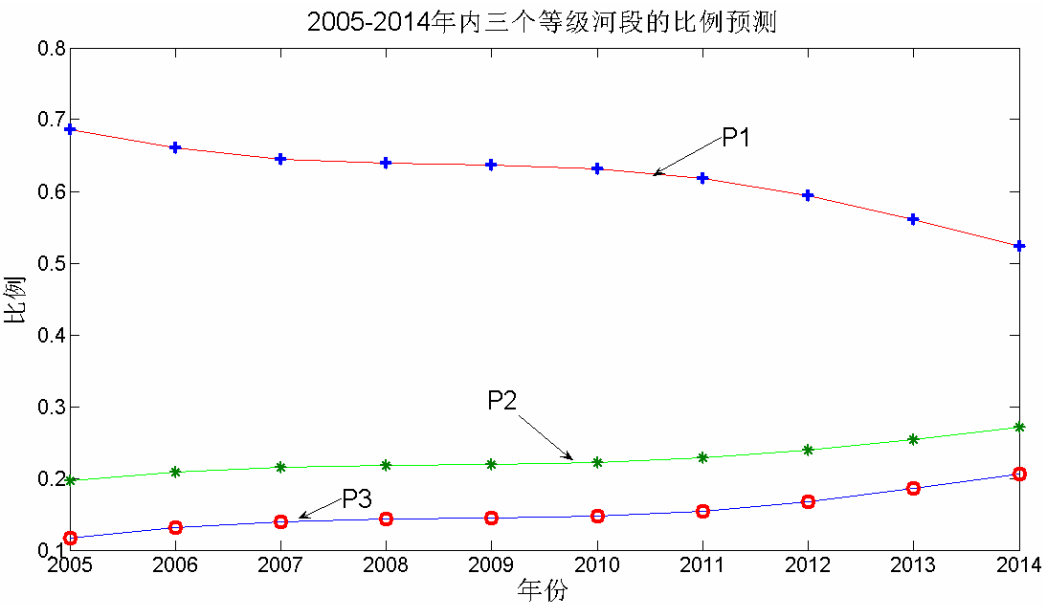
如上图，剔除 98 年大洪水与 2001 年大旱这两个异常点后，正弦曲线对剩余的数据拟合度很高。以这条正弦曲线来预测未来十年的水流量数据为(单位： 千亿立方米)：

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
水流量	9.297	9.162	9.230	9.473	9.789	10.04	10.13	10.01	9.736	9.424

8.4.5 未来十年内各等级水河段比例的趋势

结合上述污水排放量 W 与水流量 F 的预测值，我们可以得到未来十年中单位流量污染物质量 R 的值，代入预测模型，可以得到未来十年内三种水质等级对应河段所占比例的变化趋势如下表：

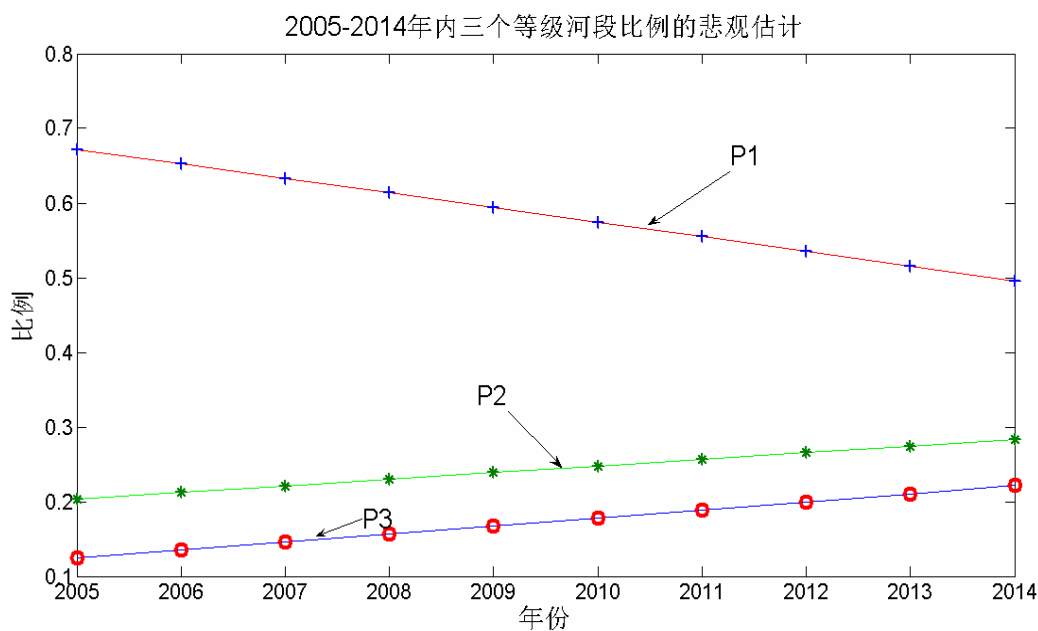
预测	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
P_1	68.5%	66.1%	64.5%	63.9%	63.6%	63.1%	61.7%	59.4%	56.1%	52.3%
P_2	19.7%	20.9%	21.6%	21.8%	22.0%	22.2%	22.8%	23.9%	25.4%	27.1%
P_3	11.7%	13.1%	13.9%	14.3%	14.4%	14.7%	15.5%	16.7%	18.6%	20.6%



分析预测数据可知，未来十年内，一等水河段所占的比例将从 68%大幅下降至将近 50%，而三等水所占的比例由 11.7%跃增到 20.6%，增幅将近 100%，二等水变化较缓，但仍然是上升的趋势。以上数据表明，在 10 年后，长江流域几万公里的河段内将有一半的河段的水质量不能达到可利用饮用水的标准，同时有超过 1/5 的河段的水处于完全不能利用的劣 V 级水。

考虑到长江流域居民用水与工业用水量的急剧增加，以及三峡水利工程对水流量的影响，未来十年水流量的变化趋势很可能是下降的。我们以 1995—2004 年 10 内的最小水流量数据作为未来 10 年的水流量，得出未来十年水质的悲观估计如下：

预测	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
P ₁	67.1%	65.2%	63.3%	61.4%	59.4%	57.5%	55.5%	53.5%	51.5%	49.5%
P ₂	20.4%	21.2%	22.1%	23.0%	23.9%	24.8%	25.6%	26.5%	27.5%	28.4%
P ₃	12.5%	13.6%	14.6%	15.7%	16.7%	17.8%	18.9%	20.0%	21.0%	22.1%



从上表可知，在悲观估计中，十年后可饮用水质的河段不到总河段长度的一半；二等水和三等水的河段长度约为总河段长度的三成和二成。此最悲观估计表征我们未来十年可能遇到的最坏情况的估计值。

两种估计方法都是对未来水质的有效估计，环境部门可以根据当前环保资金的富余程度，采纳不同的估计值拟定相应的净化对策。

9.模型应用：对问题 4 的解答

由题意，需要控制Ⅳ，Ⅴ类水质在全流域比例在20%以内，并且没有劣Ⅴ类水（具体操作时认为劣Ⅴ类水占到全流域万分之一以下），以上述条件为约束，结合线性预测模型的函数关系，以最小化未来十年处理总污水量为目标，建立线性规划模型如下：

$$\begin{cases} \text{Min.} X = \sum_{t=2005}^{2014} x(t) \\ 0.0039 + \frac{6.34 \cdot [W(t) - x(t)]}{F(t)} < 0.2 \\ \frac{7.64 \cdot [W(t) - x(t)]}{F(t)} - 0.1157 < 0.0001 \\ t = 2005, 2006, \dots, 2014 \end{cases}$$

其中 $x(t)$ 表示年份 t 处理废水的量，用 lingo 编程求得十年最少处理废水总量 1925.4 亿吨，每年处理量如下表（单位：亿吨）：

年份	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
处理量	141.5	155.8	167.1	175.8	183.5	192.2	203.5	218.1	235.1	252.8

分析 lingo 对于松弛变量和冗余变量的输出结果，我们发现劣 V 类水质的百分比限制的约束条件的冗余变量皆为 0，即约束全为紧约束；而对于 IV、V 类水质河段的比例限制的冗余变量皆大于零，即去掉这些约束对优化结果没有影响。这说明要达到题目所给的水质标准，治理劣 V 类水比治理 IV、V 类水的代价要高。

10. 污染治理方略：对问题 5 的回答

长江水资源是流域 4 亿多人口赖以生存的物质基础，也是我国南方社会经济发展的基本保障。其水污染治理关系重大，以下是我们的意见和建议：

1 控制废水排放，推行污水水资源化

从本文的分析可知：废水排放量直接影响水质的好坏，污染治理的首要任务是控制废水排放量。要控制工业废水，需充分利用价格杠杆，提高工业用水价格，推动产业结构调整，促使企业改造工业设备和生产工艺，降低用水量。同时加大政策监管力度，实行惩罚或关闭等措施，强制重污染企业对排放废水进行处理。而对于生活废水的控制，应提高城乡居民的节水意识，鼓励使用无污或少污洗涤用品。这需要持久且理性的社会舆论，以增强公众对水资源保护重要性、紧迫性的认识。

长江流域农业发达，农田化肥污染也是水污染的重要来源。应发展节水农业，将传统的粗放型高消耗、高污染、低效益农业向集约型低污染、低消耗、高效益农业转变，减少化肥农药用量，控制面源污染。

此外，污水水资源化也是治理污染的重要途径。依靠科学技术节水，实行一水多用优质优用，提高水循环利用率。

2 保护生态环境，防止水土流失与湖泊富营养化

由洞庭湖与鄱阳湖入出口水质变化可看出，湖泊对长江水质的净化有重要作用。但近年来水体富营养化进程加快，流域内湖泊富营养化问题日益突出，湖泊的调节作用被减弱。此外，长江上游水土流失严重，加剧了流域生态环境的破坏，这些都是长江水质恶化的原因。

3 科学规划，全流域统筹管理

题中所给数据显示长江部分支流水质相当之差。在长江流域内，岷江、湘江、赣江、沱江及京杭运河等支流污染都很严重。要治理长江污染，首先应当治理这些支流的污染源。同时，流域间各个地区间生态环境互相影响与牵制。对修筑水利设施，建造大型工厂等工程，应当从全流域角度统筹考虑与规划。在管理上，给予长江水利委员会对各省水利的领导权，在规划研究上，完善水环境监测网络，做好流域水功能区划和流域水资源保护规划。

4 建立和完善水资源保护法规体系

逐步建立起合理的水价体系，制定鼓励节约用水的政策，执行取水许可证制度，合理征收水资源费，落实排污总量控制方案，使水资源利用与保护有法可依。将水资源保护规划作为环境规划的组成部分，纳入国家和地方国民经济发展规划。

5 在南水北调工程中密切观察水质变化

南水北调工程将引起长江流域的生态环境一系列变化，必然将影响未来长江的水质，在工程建设中应注意继续保持引水区丹江口水库的良好水质，保护沿途生态环境。

11.模型评价

11.1 时间序列分析方法和多元分析方法

对于未来 10 年长江水质的预测，我们用到了时间序列分析方法和多元分析方法，前者通过挖掘水质变量随时间变化的规律给出预测，后者着重分析水质变量被其相关因素的作用和影响的规律。

时间序列分析中，我们采用了指平滑法和神经网络，二者的预测是基于对于历史数据的加权或者学习。指数平滑法中，预测值依赖历史数据，远期的预测要依赖其相对前期的预测结果，那么远期的预测的效果势必会有相当大的偏差，指数平滑法更适用于短期的预测。神经网络算法的数据样本过少，历史数据与新数据差异度大，使网络学习不充分。同时，其预测值也由历史数据及新生成数据决定，并因此造成了误差累积。虽然使用时间滑动窗技术可减少误差的累积，但将使训练样本数量进一步减少，在本题中并不可行。时间序列分析两种方法的预测效果均不尽人意。

多元分析法中，我们着重挖掘与水质百分比变量的相关因素，找到确定的函数关系，那么只要能够确定其相关或影响因素的未来值，就能够求出变量的预测值，则预测的效果不会因为远期而变差。

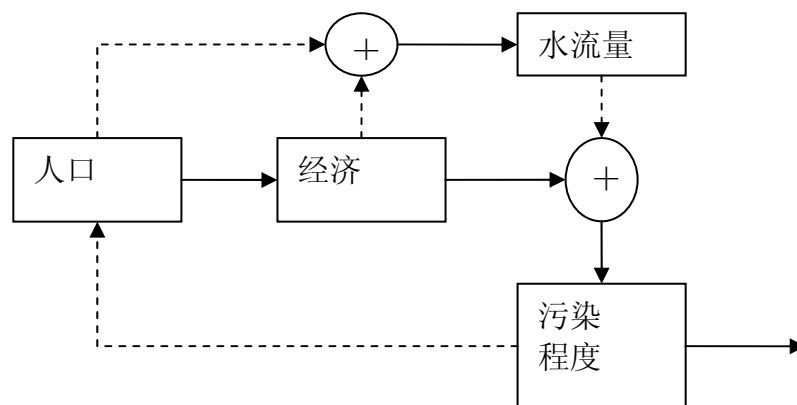
本题用 10 年的历史数据来预测未来 10 年的值，相对来讲更适合用多元分析法解决。

11.2 线性预测模型的局限性

不同水质河段的比例与单位流量污水量 R 的线性关系，只有当 R 在一定的范围内 ($0.015 \leq R \leq 0.0795$) 才成立。当 R 值极小时，其内在关系的非线性部分会发生会使实际值和线性预测值存在较大误差。比如，当 $R=0$ 时，即污水排放量为零时，三等水的比例应为零；然而用上文模型计算得出三等水河段的比例将是个负值。

12.模型拓展

我们在机理分析的预测模型中认为,水质由表征人为因素的的污染物投放量 W 与表征自然因素的水流量 F 唯一确定;同时,由于污染物投放量与流域人口数的线性关系已论证,我们也可以说水质是由流域人口数和水流量唯一确定的。在这个模型中,隐含的关系是人口数由人口自然增长规律确定,与污染程度无关。而实际上,由于水质下降会引起生态环境的急剧恶化,危及沿岸城市的饮用水安全,导致疾病肆虐等一系列问题,沿岸地区对流动人口和非常住居民的吸引力会随之下降,导致未来沿岸人口有下降的趋势或增长速度放缓。类比电路中的反馈关系可以很好的说明这一点。如下图:



图中,实线表示正相关关系,虚线表示负相关。我们把上图作为一个有向图来处理的话,则任何一条闭合路径都不是全由实线或虚线组成的,也就是说,各个因素之间存在制约关系,使当每一个因素的增长(减少)时,都存在限制因素,使其增长(减少)的趋势放缓甚至终止。比如说,但人口的增长变化引发经济的增长与水流量的减少时,随之而产生的污染程度的加剧会使人口的增长速率变缓,达到饱和点,甚至负增长(表现为长江流域的人口向外迁移)。此时,人流量的增加和污染程度的降低又会使水质有变好的趋势。也就是说,除了河水的自洁功能外,水质与人口的制约关系在一定程度上可认为是一种等效的自洁方式。如果将这种效应考虑进水质的预测中,将会使较长时期内的水质预测更符合实际。

13.参考文献

- [1] 陈红,《河流中化学污染的扩散及其数值仿真的研究》,重庆大学硕士论文,2001
- [2] 吕金虎 等,《混沌时间序列分析及其应用》,武汉:武汉大学出版社,2002
- [3] Simon Haykin,《神经网络原理》,北京:机械工业出版社,2004
- [4] 国家统计局,年度人口数据, <http://www.stats.gov.cn/tjsj/ndsj/> 2005.9.17
- [5] 姜启源,《数学模型(第三版)》,北京:高等教育出版社,2003
- [6] 刘则毅 等,《科学计算技术与 Matlab》,北京:科学出版社,2001
- [7] 余建英 等,《数据统计分析与 SPSS 应用》,北京:人民邮电出版社,2004